NATIONAL BUREAU OF STANDARDS REPORT

NBS Project | NBS Report

21 January 1955

1103-40-5140 | 3913

ANALYSIS OF PEBBLE DATA RECEIVED FROM H. F. ALBEE

Mary G. Natrella
Statistical Engineering Laboratory

NBS

U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

21 January 1955

Analysis of Pebble Data Received from H. F. Albee

1. Introduction and Summary

The data consisted of 2 samples of 150 pebbles each, classified by lithology, from each of 6 sites (HA-129, HA-132, HA-120, HA-134, HA-87-T, and HA-91-T). At two of the sites, HA-87-T and HA-91-T, an additional sample of 300 pebbles had been taken. A χ^2 test was applied to the 2 samples at each site, to test whether or not the samples differed more than might be expected on the basis of random sampling fluctuations. The computations of χ^2 are attached in Tables 1 through 6. Tables of χ^2 are available in many statistical texts, and some of these tables have been composited in ref. [1].

The value of χ^2 computed for a site is compared with the tabled value for some selected level of significance, usually Probability = .05, and for the proper "degrees of freedom". If the χ^2 computed from the samples exceeds the tabled critical value, it may be concluded that the samples differ <u>more</u> than might be expected in random sampling.

The results obtained for these 6 sites indicate that, on the whole, the 2 samples from a site are homogeneous and therefore that the sampling technique is a satisfactory one.

2. Detailed Results of Analysis

The following values of χ^2 are obtained in comparing the 2 samples at the various sites:

Site HA-129	$\chi^2 = 1.257$	(3 degrees of freedom)
HA-132	$\chi^2 = 6.366$	" " "
HA-120	$\chi^2 = 3.896$	" " "
HA-134	$\chi^2 = 1.442$	" " "
HA-91-T	$\chi^2 = 0.096$	" " "
	(for comparison of 2 samples of 150)	
HA-91-T	$\chi^2 = 3.287$	3 d.f.
	(for comparison of 2 samples of 300)	
HA-87-T	$\chi^2 = 10.685$	3 d.f.

For sites HA-129, HA-132, HA-120, and HA-134, χ^2 does not exceed $\chi^2_{.05}$ and therefore the 2 samples do not differ excessively one from the other.

At HA-91-T, the comparison of 2 samples of 150 yields a $\chi^2 = 0.096$, which is so small (the probability of a larger value of χ^2 exceeds .99) that it might lead to a suspicion that these 2 samples were not random. In other words, these samples agree too well, which may be due to unusual luck or to a defect in the sampling technique. Since there was an additional sample of 300 available for HA-91-T, the 2 samples of 150 were pooled for comparison with it, and the resulting χ^2 (=3.287) is not significant and shows good agreement between the 2 samples of 300.

At HA-87-T, the samples of 150 show larger differences than would be expected. The large value of χ^2 (=10.685) apparently comes from the relatively large discrepancy in the "other" classification. In view of the non-homogeneity these samples were not pooled for comparison with the sample of 300. If the comparison of samples of 300 could be made, however, it would undoubtedly indicate lack of homogeneity due to the quartz class.

NOTE: For HA-91-T and HA-87-T, the small classes have been merged into one "other" classification. This was done not only to achieve uniformity with the other 4 sites, but because of certain theoretical difficulties in applying the χ^2 test to classes with very small counts (less than 10 as a minimum).

3. Short-cut Methods of Computing χ^2

3a.

The usual definition of χ^2, originally given by Karl Pearson [2], is:

$$\chi^2 = \Sigma \frac{(\text{observed frequency} - \text{expected frequency})^2}{\text{expected frequency}}$$

3b.

A simplified formula, given by R. A. Fisher [3], for χ^2 in comparing 2 observed frequency distributions is:

$$\chi^2 = \sum_{i=1}^{s} \left[\frac{N N'}{f_i + f'_i} \left(\frac{f_i}{N} - \frac{f'_i}{N'} \right)^2 \right]$$

where the class frequencies are f_i and f_i' respectively for the ith class, N and N' are the total counts in each sample, and s is the number of classes. The degrees of freedom for χ^2 is s - 1.

3c.

For two observed distributions with equal total counts (as in this case), the formula in 3b above simplifies to:

$$\chi^2 = \sum_{i=1}^{s} \left\{ \frac{(f_i - f_i')^2}{f_i + f_i'} \right\} \qquad \text{d.f.} = s - 1$$

In other words, (1) for each class, take the difference between the 2 counts, square this difference and divide by the sum of the 2 counts (2) Sum these quantities over all classes. This form of computation is somewhat simpler than the usual one, and has the advantage that it does not require taking deviations from the average count and therefore eliminates resulting questions as to how many decimal places to carry.

3d.

For three observed distributions with equal total counts, there is a similar short-cut. (1) For each class, take all 3 possible differences among the counts; take the sum of the squares of the 3 differences and divide by the sum of the 3 counts (2) Sum over all classes.

Mary G. Natrella
Statistical Engineering Laboratory
National Bureau of Standards
21 January 1955

REFERENCES

[1] National Bureau of Standards Report, NAML 51-2 "Some Percentage Points of the χ^2 Distribution", August 1950.

[2] K. Pearson, "On a Criterion that a Given System of Deviations from the Probable in the Case of a Correlated System of Variables is Such that It Can Be Reasonably Supposed to Have Arisen from Random Sampling", Phil. Mag. Series V, 50, 1900, 157-175.

[3] R. A. Fisher, "On the Interpretation of χ^2 from Contingency Tables and the Calculation of P", J.R.S.S. 85, 1922, 87-94.

Table 1

"C Group
Interriver Area, Utah

Samples No. HA-129-A
HA-129-B

Lithology Class	Frequency: Sample A	Frequency: Sample B	A+B	A-B	$\frac{(A-B)^2}{A+B}$
Quartzite	49	42	91	7	0.538
Quartz	11	10	21	1	0.048
Chert	78	82	160	-4	0.100
Other	12	16	28	-4	0.571
					χ^2 =
TOTAL	150	150	300	0	1.257
					d.f. = 3

TABLE 2

So. Bridger Jack
Elk Ridge Area, Utah

Samples No. HA-132-A
HA-132-B

Lithology Class	Frequency Sample A	Frequency Sample B	A+B	A-B	$\frac{(A-B)^2}{A+B}$
Quartzite	30	24	54	6	0.667
Quartz	13	5	18	8	3.556
Chert	104	120	224	-16	1.143
Other	3	1	4	2	1.000
TOTAL	150	150	300	0	$\chi^2 =$ 6.366

d.f. = 3

TABLE 3

Middle Trail
Green River Desert Area, Utah

Samples No. KA-120-A
KA-120-B

Lithology Class	Frequency Sample A	Frequency Sample B	A+B	A-B	$\frac{(A-B)^2}{A+B}$
Quartzite	75	60	135	15	1.667
Quartz	16	19	35	-3	0.257
Chert	49	63	112	-14	1.750
Other	10	8	18	2	0.222
					χ^2 =
TOTAL	150	150	300	0	3.896
					d.f. = 3

TABLE 4

The Notch
Elk Ridge Area, Utah

Samples No. HA-134-A
HA-134-B

Lithology Class	Frequency Sample A	Sample B	A+B	A-B	$\frac{(A-B)^2}{A+B}$
Quartzite	60	69	129	-9	0.628
Quartz	32	26	58	6	0.621
Chert	46	45	91	1	0.011
Other	12	10	22	2	0.182
					χ^2 =
TOTAL	150	150	300	0	1.442
					d.f. = 3

TABLE 8c

Buckhorn Draw
Emery Co., Utah

HA-91-T
(Samples of 150)

Lithology Class		Frequency Sample B_1	Frequency Sample B_2	B_1+B_2	B_1-B_2	$\frac{(B_1-B_2)^2}{B_1+B_2}$
Quartzite		93	94	187	-1	0.005
Quartz		19	20	39	-1	0.026
Chert		32	30	62	2	0.065
Sil. ls.	"Other"	6	6			
Limestone	"Other"	—	—			
Claystone	"Other"	—	—			
Sandstone	"Other"	—	—			
"Other" (total)		6	6	12	0	0
						$\chi^2 =$
TOTAL		150	150	300	0	0.096
						d.f. = 3

TABLE 5b

Buckhorn Draw
Emery Co., Utah

HA-91-T
(Samples of 300)

Lithology Class		Frequency Sample A		Frequency Sample B		A+B	A-B	$\frac{(A-B)^2}{A+B}$
Quartzite		188		187		375	1	0.003
Quartz		48		39		87	9	0.931
Chert		48		62		110	-14	1.782
Sil. ls.	"Other"	11	16	12	12	28	4	0.571
Limestone		2		--				
Claystone		1		--				
Sandstone		2		--				
TOTAL		300		300		600	0	$X^2 =$ 3.287
								d.f. = 3

TABLE 6

Moccasin, Arizona

HA-87-T
(Samples of 150)

Lithology Class		Frequency Sample B_1		Sample B_2		B_1+B_2	B_1-B_2	$\frac{(B_1-B_2)^2}{B_1+B_2}$
Quartzite		92		112		204	-20	1.961
Quartz		17		17		34	0	0
Chert		25		17		42	8	1.524
Claystone		8		1				
Siltstone		4		2				
Sandstone	"Other"	2	16	--	4	20	12	7.200
Sil. ls.		1		1				
Limestone		1		--				
TOTAL		150		150		300	0	$\chi^2 =$ 10.685

d.f. = 3